

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Masure, H. Robert
Rosenow, Carsten I.
Tuomanen, Elaine
Wizemann, Theresa M.

(ii) TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
ANTI-PNEUMOCOCCAL VACCINES

(iii) NUMBER OF SEQUENCES: 25

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: David A. Jackson, Esq.
(B) STREET: 411 Hackensack Ave, Continental Plaza, 4th
Floor
(C) CITY: Hackensack
(D) STATE: New Jersey
(E) COUNTRY: USA
(F) ZIP: 07601

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Jackson Esq., David A.
(B) REGISTRATION NUMBER: 26,742
(C) REFERENCE/DOCKET NUMBER: 600-1-158 ..

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 201-487-5800
- (B) TELEFAX: 201-343-1684

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP112

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Xaa	Glu	Asn	Glu	Gly	Ser	Thr	Gln	Ala	Ala	Thr	Ser	Ser	Asn	Met	Ala
1														10	

Lys	Thr	Glu	His	Arg	Lys	Ala	Ala	Lys	Gln	Val	Val	Asp	Glu			
														20	25	30

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP90

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Arg Glu Phe Ser Leu Glu Lys Thr Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP84

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Xaa Arg Glu Phe Ser Leu Glu Lys Thr Arg Asn Ile Gly Ile Met Ala
1 5 10 15

His Val Asp Ala Gly Lys Thr
20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(A) DESCRIPTION: cBP80

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Xaa Lys Xaa Xaa Trp Gln Xaa Lys Gln Tyr Leu Lys Glu Asp Gly Ser
1 5 10 15

Gln Ala Ala Asn Glu Xaa Val Phe Asp Thr Ala
20 25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(A) DESCRIPTION: cBP78

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gln Lys Ile Ile Gly Ile Asp Leu Gly Thr Thr Asn Ser Ala Val Ala
1 5 10 15

Val Leu Glu Gly Thr Glu Ser Lys Ile Ile Ala Asn Pro Glu
20 25 30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP70

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Xaa	Xaa	Xaa	Glu	Val	Ala	Lys	Xaa	Ser	Gln	Asp	Thr	Thr	Thr	Ala	Ser
1				5							10				15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP60

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Xaa	Asn	Glu	Arg	Val	Lys	Ile	Val	Ala	Thr	Leu	Gly	Pro	Ala	Val	Glu
1				5					10						15

Gly Arg Gly

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP50 pep

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Xaa Ile Ile Xaa Xaa Val Tyr Ala Arg Glu Val Leu Asp Ser Arg Gly
1 5 10 15

Asn Pro

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP112-Int1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Glu Asp Arg Arg Asn Tyr Pro Thr Asn Thr Tyr Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP112-Int2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Xaa Asp Asp Gln Gln Ala Glu Glu Asp Tyr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Trp Gln Pro Pro Arg Ala Arg Ile
1 5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: 50 KDa protein

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *S. pneumoniae*

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ile Thr Asp Val Tyr Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(A) DESCRIPTION: 50 KDa protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ile Val Asp Val Arg Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: S. pneumoniae

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCCCGCGGCA ACCAACACT TGAAGTAGAA GTTTACACTG AATCAGGTGC TTTCGGACGT

GGTATGGTTC CATCAGGAGC TTCTACTGGT GAACACGAAG CAGTTGAAC TCGCGACGGT	120
GACAAATCTC GTTACGGTGG TCTTGGTACA CAAAAAGCTG TTGACAACGT AAACAACATC	180
ATTGCTGAGG CCATCATTGG CTACGATGTA	210

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 210 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *B. subtilis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCCCGCGGCA ACCAACAGT TGAAGTTGAA GTATATACAG AAACAGGAGC TTTCGGCCGC	60
GCATTAGTGC CAAGCGGAGC TTCTACAGGT CAATACGAAG CGGTTGAGCT TCGTGACGGC	120
GACAAAGACC GTTACCTTGG AAAAGGCGTG TTAACAGCTG TTAACAACGT AAACGAAATC	180
ATTGCTCCAG AGCTTCTTGG CTTTGATGTA	210

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 amino acids
 - (B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ser Arg Gly Asn Pro Thr Leu Glu Val Glu Val Tyr Thr Glu Ser Gly
1 5 10 15

Ala Phe Gly Arg Gly Met Val Pro Ser Gly Ala Ser Thr Gly Glu His
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Ser Arg Tyr Gly Gly Leu
35 40 45

Gly Thr Gln Lys Ala Val Asp Asn Val Asn Asn Ile Ile Ala Glu Ala
50 55 60

Ile Ile Gly Tyr Asp Val
65 70

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ser Arg Gly Asn Pro Thr Val Glu Val Glu Val Tyr Thr Glu Thr Gly
1 5 10 15

Ala Phe Gly Arg Ala Leu Val Pro Ser Gly Ala Ser Thr Gly Gln Tyr
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Asp Arg Tyr Leu Gly Lys
35 40 45

Gly Val Leu Thr Ala Val Asn Asn Val Asn Glu Ile Ile Ala Pro Glu
50 55 60

Leu Leu Gly Phe Asp Val
65 70

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1511 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCCCCGGCA ACCCAACACT TGAAGTAGAA GTTTACACTG AATCAGGTGC TTTCGGACGT	60
GGTATGGTTC CATCAGGAGC TTCTACTGGT GAACACGAAG CAGTTGAAC TCGCGACGGT	120
GACAAATCTC GTTACGGTGG TCTTGGTACA CAAAAAGCTG TTGACAACGT AAACAACATC	180
ATTGCTGATT CTATCATTGG CTACGATGTA CGTGATCAAC AAGCTATTGA CCGTGCTATG	240
ATCGCACTTG ACGGTACTCC TAACAAAGGT AAATTGGGTG CGAATGCAAT CCTCGGTGTG	300
TCTATCGCTG TAGCTCGTGC TGCTGCTGAC TACCTGAAA TCCCACTTA CAGCTACCTT	360
GGTGGATTCA ACACCAAAGT TCTTCCAACT CCAATGATGA ACATCATCAA CGGTGGTTCT	420
CACTCTGACG CTCCAATCGC TTTCCAAGAG TTCATGATCT TGCCAGTGTG TGGCCAAACA	480
TTTAAAGAAA CCCTTCGTTA CGGTGCTGAA ATCTTCCACG CTCTTAAGAA AATCCTTAAA	540
TCACGTGGTT TGGAAACTGC CGTAGGTGAC GAAGGTGGAT TCGCTCCTCG TTTCGAAGGA	600
ACTGAAGATG GTGTTGAAAC TATCCTTGCT GCGATTGAAG CTGCTGGATA TGTACCAGGT	660
AAAGACGTAT TTCTCGGATT TGACTGTGCT TCATCAGAAT TCTACGATAA AGAACGTAAA	720
GTTTACGACT ACACCAAATT TGAAGGCGAA GGTGCTGCTG TTCTGACATC TGCAGAACAA	780
ATCGACTACC TTGAAGAATT GGTTAACAAA TACCCAATCA TCACTATTGA AGATGGTATG	840
GATGAAAACG ACTGGGATGG TTGGAAAGCT CTTACTGAAC GTCTTGGTAA GAAAGTACAA	900
CTTGTGGTG ACGACTTCTT CGTAACAAAC ACTGACTACC TTGCACGTGG TATCCAAAAA	960
GGTGTGCTA ACTCAATCCT TATCAAAGTT AACCAAATCG GTACTCTTAC TGAAACTTTT	1020
GAAGCTATCG AAATGGCTAA AGAAGCTGGT TACACTGCTG TTGTATCACA CCGTTCAGGT	1080

GAAACTGAAG ATTCAACAAT CGCTGATATC GCGATTGCAA CTAACGCAGG ACAAAATCAAG	1140
ACTGGTTCAC TTTCACGTAC AGACCGTATC GCTAAATACA ACCAATTGCT TCGTATCGAA	1200
GACCAACTTG GTGAAGTAGC AGAATATCGT GGATTGAAAT CATTCTACGA CCTTAAAAAA	1260
TAAAAATAGTA CAGTGAACTA TTTTATCCCC GAACCATGAA ATTCAAATTC CGGCCCTTGA	1320
TTTAACAACT TTCTAGCCCC CCGATTATC CGAGGGGATA TTTTGTTTT AAAGGGCAAA	1380
AAGGACAAAC TTTTAAAAT CTTTCATAC ATTACCATAC TTGATGCTCC CCTCCTTCTT	1440
GTACTGGTTT AAGAATATAC TCCATCGATG ACACCGTCGG ATGTTAGGA TTATTCGACA	1500
ACAAGTTGAA T	1511

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 420 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ser	Arg	Gly	Asn	Pro	Thr	Leu	Glu	Val	Glu	Val	Tyr	Thr	Glu	Ser	Gly
1															

Ala	Phe	Gly	Arg	Gly	Met	Val	Pro	Ser	Gly	Ala	Ser	Thr	Gly	Glu	His
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

20

25

30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Ser Arg Tyr Gly Gly Leu
35 40 45

Gly Thr Gln Lys Ala Val Asp Asn Val Asn Asn Ile Ile Ala Asp Ser
50 55 60

Ile Ile Gly Tyr Asp Val Arg Asp Gln Gln Ala Ile Asp Arg Ala Met
65 70 75 80

Ile Ala Leu Asp Gly Thr Pro Asn Lys Gly Lys Leu Gly Ala Asn Ala
85 90 95

Ile Leu Gly Val Ser Ile Ala Val Ala Arg Ala Ala Asp Tyr Leu
100 105 110

Glu Ile Pro Leu Tyr Ser Tyr Leu Gly Gly Phe Asn Thr Lys Val Leu
115 120 125

Pro Thr Pro Met Met Asn Ile Ile Asn Gly Gly Ser His Ser Asp Ala
130 135 140

Pro Ile Ala Phe Gln Glu Phe Met Ile Leu Pro Val Gly Ala Pro Thr
145 150 155 160

Phe Lys Glu Thr Leu Arg Tyr Gly Ala Glu Ile Phe His Ala Leu Lys
165 170 175

Lys Ile Leu Lys Ser Arg Gly Leu Glu Thr Ala Val Gly Asp Glu Gly
180 185 190

Gly Phe Ala Pro Arg Phe Glu Gly Thr Glu Asp Gly Val Glu Thr Ile
195 200 205

Leu Ala Ala Ile Glu Ala Ala Gly Tyr Val Pro Gly Lys Asp Val Phe
210 215 220

Leu Gly Phe Asp Cys Ala Ser Ser Glu Phe Tyr Asp Lys Glu Arg Lys
225 230 235 240

Val	Tyr	Asp	Tyr	Thr	Lys	Phe	Glu	Gly	Glu	Gly	Ala	Ala	Val	Arg	Thr
245															255
Ser Ala Glu Gln Ile Asp Tyr Leu Glu Glu Leu Val Asn Lys Tyr Pro															
260									265						270
Ile Ile Thr Ile Glu Asp Gly Met Asp Glu Asn Asp Trp Asp Gly Trp															
275							280								285
Lys Ala Leu Thr Glu Arg Leu Gly Lys Lys Val Gln Leu Val Gly Asp															
290					295									300	
Asp Phe Phe Val Thr Asn Thr Asp Tyr Leu Ala Arg Gly Ile Gln Lys															
305				310					315						320
Gly Ala Ala Asn Ser Ile Leu Ile Lys Val Asn Gln Ile Gly Thr Leu															
325					330										335
Thr Glu Thr Phe Glu Ala Ile Glu Met Ala Lys Glu Ala Gly Tyr Thr															
340					345										350
Ala Val Val Ser His Arg Ser Gly Glu Thr Glu Asp Ser Thr Ile Ala															
355					360									365	
Asp Ile Ala Ile Ala Thr Asn Ala Gly Gln Ile Lys Thr Gly Ser Leu															
370					375									380	
Ser Arg Thr Asp Arg Ile Ala Lys Tyr Asn Gln Leu Leu Arg Ile Glu															
385					390				395						400
Asp Gln Leu Gly Glu Val Ala Glu Tyr Arg Gly Leu Lys Ser Phe Tyr															
405							410								415
Asp Leu Lys Lys															
420															

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 429 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AAAAAAAGGT	GCAGAACGCTG	AGAAGAAGGT	TGAAGAAGCT	GAGAAAAAAG	CCAAGGATCA	60
AAAAGAAGAA	GATGCCGTA	ACTACCCAAC	CAATACTTAC	AAAACGCTTG	ACCTTGAAAT	120
TGCTGAGTCC	GATGTGAAAG	TTAAAGAAGC	GGAGCTTGAA	CTAGTAAAAG	AGGAAGCTAA	180
GGAACCTCGA	GACGAGGAAA	AAATTAAGCA	AGCAAAAGCG	AAAGTTGAGA	GTAAAAAAGC	240
TGAGGCTACA	AGGTTAGAAA	ACATCAAGAC	AGATCGTAAA	AAAGCAGAAG	AAGAAGCTAA	300
ACGAAAAGCA	GCAGAAGAAG	ATAAAGTTAA	AGAAAAACCA	GCTGAACAAC	CACAACCAGC	360
GCCGGTTACT	CAACCAGAAA	AACCAGCTCC	AAAACCAGAG	AAGCCAGCTG	AACAACCAAA	420
AGCAGAAAA						429

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Lys Lys Val Ala Glu Ala Glu Lys Lys Val Glu Glu Ala Glu Lys Lys
1 5 10 15

Ala Lys Asp Gln Lys Glu Glu Asp Arg Arg Asn Tyr Pro Thr Asn Thr
20 25 30

Tyr Lys Thr Leu Asp Leu Glu Ile Ala Glu Ser Asp Val Lys Val Lys
35 40 45

Glu Ala Glu Leu Glu Leu Val Lys Glu Glu Ala Lys Glu Pro Arg Asp
50 55 60

Glu Glu Lys Ile Lys Gln Ala Lys Ala Lys Val Glu Ser Lys Lys Ala
65 70 75 80

Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp Arg Lys Lys Ala Glu
85 90 95

Glu Glu Ala Lys Arg Lys Ala Ala Glu Glu Asp Lys Val Lys Glu Lys
100 105 110

Pro Ala Glu Gln Pro Gln Pro Ala Pro Val Thr Gln Pro Glu Lys Pro
115 120 125

Ala Pro Lys Pro Glu Lys Pro Ala Glu Gln Pro Lys Ala Glu
130 135 140

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTCTTNCTC GATGTCTCNG TNGCCAT

27

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AGCATAAGTCT TCTTCGACTT GTTGATCATC

30

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2462 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCAAGCAGCC ACTTCTTCTA ATATGGCAAA GACAGAACAT AGGAAAGCTG CTAACAAGC	60
GTCGATGAAT ATATAGAAAA AATGATGAGG GAGATTCAAC TAGATAGAAG AAAACATACC	120
CAAAATGTCG CCTTAAACAT AAAGTTGAGC GCAATTAAAA CGAAGTATTG GCGTGAATTA	180
AATGTTTAG AAGAGAAGTC GAAAGATGAG TTGCCGTCAG AAATAAAAGC AAAGTTAGAC	240
GCAGCTTTG AGAAGTTAA AAAAGATACA TTGAAACCAG GAGAAAAGGT AGCAGAAGCT	300
AAGAAGAAGG TTGAAGAAGC TAAGAAAAAA GCCGAGGATC AAAAAGAAGA AGATCGTCGT	360
AACTACCAA CCAATACTTA CAAAACGCTT GAACTTGAAA TTGCTGAGTT CGATGTGAAA	420
GTTAAAGAAG CGGAGCTTGA ACTAGTAAAA GAGGAAGCTA AAGAATCTCG AAACGAGGGC	480
ACAATTAAGC AAGCAAAAGA GAAAGTTGAG AGTAAAAAAG CTGAGGCTAC AAGGTTAGAA	540
AACATCAAGA CAGATCGTAA AAAAGCAGAA GAAGAACGCTA AACGAAAAGC AGATGCTAAG	600
TTGAAGGAAG CTAATGTAGC GACTTCAGAT CAAGGTAAAC CAAAGGGCG GGCAAAACGA	660
GGAGTTCCCTG GAGAGCTAGC AACACCTGAT AAAAAAGAAA ATGATGCGAA GTCTTCAGAT	720

TCTAGCGTAG	GTGAAGAAC	TCTTCCAAGC	TCATCCCTGA	AATCAGGAAA	AAAGGTAGCA	780
GAAGCTGAGA	AGAAGGTTGA	AGAAGCTGAG	AAAAAAGCCA	AGGATCAAAA	AGAAGAAGAT	840
CGCCGTAACT	ACCCAACCAA	TACTTACAAA	ACGCTTGACC	TTGAAATTGC	TGAGTCCGAT	900
GTGAAAGTTA	AAGAAGCGGA	GCTTGAACTA	GTAAAAGAGG	AAGCTAAGGA	ACCTCGAGAC	960
GAGGAAAAAA	TTAAGCAAGC	AAAAGCGAAA	GTTGAGAGTA	AAAAAGCTGA	GGCTACAAGG	1020
TTAGAAAACA	TCAAGACAGA	TCGTAAAAAA	GCAGAAGAAG	AAGCTAAACG	AAAAGCAGCA	1080
GAAGAAGATA	AAGTTAAAGA	AAAACCAGCT	GAACAACCAC	AACCAGCGCC	GGCTACTCAA	1140
CCAGAAAAAC	CAGCTCCAAA	ACCAAGAGAAG	CCAGCTGAAC	AACCAAAAGC	AGAAAAAAACA	1200
GATGATCAAC	AAGCTGAAGA	AGACTATGCT	CGTAGATCAG	AAGAAGAATA	TAATCGCTTG	1260
ACTCAACAGC	AACCGCCAAA	AACTGAAAAA	CCAGCACAAC	CATCTACTCC	AAAAACAGGC	1320
TGGAAACAAG	AAAACGGTAT	GTGGTACTTC	TACAATACTG	ATGGTTCAAT	GGCAACAGGA	1380
TGGCTCCAAA	ACAACGGTTC	ATGGTACTAT	CTAAACGCTA	ATGGTGCTAT	GGCGACAGGA	1440
TGGCTCCAAA	ACAATGGTTC	ATGGTACTAT	CTAAACGCTA	ATGGTTCAAT	GGCAACAGGA	1500
TGGCTCCAAA	ACAATGGTTC	ATGGTACTAC	CTAAACGCTA	ATGGTGCTAT	GGCGACAGGA	1560
TGGCTCCAAT	ACAATGGTTC	ATGGTACTAC	CTAAACAGCA	ATGGCGCTAT	GGCGACAGGA	1620
TGGCTCCAAT	ACAATGGCTC	ATGGTACTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGA	1680
TGGCTCCAAA	ACAACGGTTC	ATGGTACTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGA	1740
TGGCTCCAAT	ACAACGGTTC	ATGGTATTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGT	1800
TGGGTGAAAG	ATGGAGATAC	CTGGTACTAT	CTTGAAGCAT	CAGGTGCTAT	GAAAGCAAGC	1860
CAATGGTTCA	AAGTATCAGA	TAAATGGTAC	TATGTCAATG	GCTCAGGTGC	CCTTGCAGTC	1920
AACACAACTG	TAGATGGCTA	TGGAGTCAAT	GCCAATGGTG	AATGGGTAAA	CTAACCTAA	1980

TATAACTAGT TAATACTGAC TTCCTGTAAG AACTTTTAA AGTATTCCCT ACAAATACCA	2040
TATCCTTCA GTAGATAATA TACCCTGTA GGAAGTTAG ATTAAAAAAT AACTCTGTAA	2100
TCTCTAGCCG GATTTATAGC GCTAGAGACT ACGGAGTTT TTTGATGAGG AAAGAATGGC	2160
GGCATTCAAG AGACTCTTA AGAGAGTTAC GGGTTTAAA CTATTAAGCC TTCTCCAATT	2220
GCAAGAGGCT TCAATCTCTG CTAGGGTGCT AGCTTGCAG ATGGCTCCAC GGAGTTGGC	2280
AGCGCCAGAT GTTCCACCGA GATAGTGAGG AGCGAGGCCG CGGAATTCAC GAACTGCGAC	2340
GTGTTCTCCT TTGAGGTTAA TCAATCGTTT CAAAGTTCG TAGGCGATCT TCATCTTGTC	2400
TTCAAAGGTC AAATCAGGTA GGATTCTCC TGTTCAAAG TTTATGGTGG CCCTGGTTGA	2460
AG	2462

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 631 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met	Met	Arg	Glu	Ile	Gln	Leu	Asp	Arg	Arg	Lys	His	Thr	Gln	Asn	Val
1															

Ala	Leu	Asn	Ile	Lys	Leu	Ser	Ala	Ile	Lys	Thr	Lys	Tyr	Leu	Arg	Glu
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

20

25

30

Leu Asn Val Leu Glu Glu Lys Ser Lys Asp Glu Leu Pro Ser Glu Ile
35 40 45

Lys Ala Lys Leu Asp Ala Ala Phe Glu Lys Phe Lys Lys Asp Thr Leu
50 55 60

Lys Pro Gly Glu Lys Val Ala Glu Ala Lys Lys Lys Val Glu Glu Ala
65 70 75 80

Lys Lys Lys Ala Glu Asp Gln Lys Glu Glu Asp Arg Arg Asn Tyr Pro
85 90 95

Thr Asn Thr Tyr Lys Thr Leu Glu Leu Glu Ile Ala Glu Phe Asp Val
100 105 110

Lys Val Lys Glu Ala Glu Leu Glu Leu Val Lys Glu Glu Ala Lys Glu
115 120 125

Ser Arg Asn Glu Gly Thr Ile Lys Gln Ala Lys Glu Lys Val Glu Ser
130 135 140

Lys Lys Ala Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp Arg Lys
145 150 155 160

Lys Ala Glu Glu Ala Lys Arg Lys Ala Asp Ala Lys Leu Lys Glu
165 170 175

Ala Asn Val Ala Thr Ser Asp Gln Gly Lys Pro Lys Gly Arg Ala Lys
180 185 190

Arg Gly Val Pro Gly Glu Leu Ala Thr Pro Asp Lys Lys Glu Asn Asp
195 200 205

Ala Lys Ser Ser Asp Ser Ser Val Gly Glu Glu Thr Leu Pro Ser Ser
210 215 220

Ser Leu Lys Ser Gly Lys Lys Val Ala Glu Ala Glu Lys Lys Val Glu
225 230 235 240

Glu Ala Glu Lys Lys Ala Lys Asp Gln Lys Glu Glu Asp Arg Arg Asn
245 250 255

Tyr Pro Thr Asn Thr Tyr Lys Thr Leu Asp Leu Glu Ile Ala Glu Ser
260 265 270

Asp Val Lys Val Lys Glu Ala Glu Leu Glu Leu Val Lys Glu Glu Ala
275 280 285

Lys Glu Pro Arg Asp Glu Glu Lys Ile Lys Gln Ala Lys Ala Lys Val
290 295 300

Glu Ser Lys Ala Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp
305 310 315 320

Arg Lys Lys Ala Glu Glu Ala Lys Arg Lys Ala Ala Glu Glu Asp
325 330 335

Lys Val Lys Glu Lys Pro Ala Glu Gln Pro Gln Pro Ala Pro Ala Thr
340 345 350

Gln Pro Glu Lys Pro Ala Pro Lys Pro Glu Lys Pro Ala Glu Gln Pro
355 360 365

Lys Ala Glu Lys Thr Asp Asp Gln Gln Ala Glu Glu Asp Tyr Ala Arg
370 375 380

Arg Ser Glu Glu Glu Tyr Asn Arg Leu Thr Gln Gln Gln Pro Pro Lys
385 390 395 400

Thr Glu Lys Pro Ala Gln Pro Ser Thr Pro Lys Thr Gly Trp Lys Gln
405 410 415

Glu Asn Gly Met Trp Tyr Phe Tyr Asn Thr Asp Gly Ser Met Ala Thr
420 425 430

Gly Trp Leu Gln Asn Asn Gly Ser Trp Tyr Tyr Leu Asn Ala Asn Gly
435 440 445

Ala Met Ala Thr Gly Trp Leu Gln Asn Asn Gly Ser Trp Tyr Tyr Leu
450 455 460

Asn Ala Asn Gly Ser Met Ala Thr Gly Trp Leu Gln Asn Asn Gly Ser
465 470 475 480

Trp Tyr Tyr Leu Asn Ala Asn Gly Ala Met Ala Thr Gly Trp Leu Gln
485 490 495

Tyr Asn Gly Ser Trp Tyr Tyr Leu Asn Ser Asn Gly Ala Met Ala Thr
500 505 510

Gly Trp Leu Gln Tyr Asn Gly Ser Trp Tyr Tyr Leu Asn Ala Asn Gly
515 520 525

Asp Met Ala Thr Gly Trp Leu Gln Asn Asn Gly Ser Trp Tyr Tyr Leu
530 535 540

Asn Ala Asn Gly Asp Met Ala Thr Gly Trp Leu Gln Tyr Asn Gly Ser
545 550 555 560

Trp Tyr Tyr Leu Asn Ala Asn Gly Asp Met Ala Thr Gly Trp Val Lys
565 570 575

Asp Gly Asp Thr Trp Tyr Tyr Leu Glu Ala Ser Gly Ala Met Lys Ala
580 585 590

Ser Gln Trp Phe Lys Val Ser Asp Lys Trp Tyr Tyr Val Asn Gly Ser
595 600 605

Gly Ala Leu Ala Val Asn Thr Thr Val Asp Gly Tyr Gly Val Asn Ala
610 615 620

Asn Gly Glu Trp Val Asn Glx
625 630

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Masure, H. Robert
Rosenow, Carsten I.
Tuomanen, Elaine
Wizemann, Theresa M.

(ii) TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
ANTI-PNEUMOCOCCAL VACCINES

(iii) NUMBER OF SEQUENCES: 25

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: David A. Jackson, Esq.
(B) STREET: 411 Hackensack Ave, Continental Plaza, 4th
Floor
(C) CITY: Hackensack
(D) STATE: New Jersey
(E) COUNTRY: USA
(F) ZIP: 07601

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Jackson Esq., David A.

(B) REGISTRATION NUMBER: 26,742
(C) REFERENCE/DOCKET NUMBER: 600-1-158 ..

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 201-487-5800
(B) TELEFAX: 201-343-1684

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP112

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Xaa Glu Asn Glu Gly Ser Thr Gln Ala Ala Thr Ser Ser Asn Met Ala
1 5 10 15

Lys Thr Glu His Arg Lys Ala Ala Lys Gln Val Val Asp Glu
20 25 30

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP90

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Arg Glu Phe Ser Leu Glu Lys Thr Arg

1 5 10

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP84

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Xaa Arg Glu Phe Ser Leu Glu Lys Thr Arg Asn Ile Gly Ile Met Ala

1 5 10 15

His Val Asp Ala Gly Lys Thr

20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP80

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Xaa Lys Xaa Xaa Trp Gln Xaa Lys Gln Tyr Leu Lys Glu Asp Gly Ser

1

5

10

15

Gln Ala Ala Asn Glu Xaa Val Phe Asp Thr Ala

20

25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP78

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gln Lys Ile Ile Gly Ile Asp Leu Gly Thr Thr Asn Ser Ala Val Ala
1 5 10 15

Val Leu Glu Gly Thr Glu Ser Lys Ile Ile Ala Asn Pro Glu
20 25 30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP70

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Xaa Xaa Xaa Glu Val Ala Lys Xaa Ser Gln Asp Thr Thr Ala Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP60

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Xaa	Asn	Glu	Arg	Val	Lys	Ile	Val	Ala	Thr	Leu	Gly	Pro	Ala	Val	Glu
1															15
Gly Arg Gly															

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP50 pep

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Xaa Ile Ile Xaa Xaa Val Tyr Ala Arg Glu Val Leu Asp Ser Arg Gly
1 5 10 15

Asn Pro

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP112-Int1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Glu Asp Arg Arg Asn Tyr Pro Thr Asn Thr Tyr Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP112-Int2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Xaa Asp Asp Gln Gln Ala Glu Glu Asp Tyr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Trp Gln Pro Pro Arg Ala Arg Ile
1 5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: 50 KDa protein

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: S. pneumoniae

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ile Thr Asp Val Tyr Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro

1

5

10

15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: 50 KDa protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *B. subtilis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ile Val Asp Val Arg Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCCCGCGGCA ACCCAACACT TGAAGTAGAA GTTTACACTG AATCAGGTGC TTTCGGACGT

GGTATGGTTC CATCAGGAGC TTCTACTGGT GAACACGAAG CAGTTGAAC TCGCGACGGT	120
GACAAATCTC GTTACGGTGG TCTTGGTACA CAAAAAGCTG TTGACAACGT AAACAAACATC	180
ATTGCTGAGG CCATCATTGG CTACGATGTA	210

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCCCGCGGCA ACCCAACAGT TGAAGTTGAA GTATATACAG AAACAGGAGC TTTCGGCCGC	60
GCATTAGTGC CAAGCGGAGC TTCTACAGGT CAATACGAAG CGGTTGAGCT TCGTGACGGC	120
GACAAAGACC GTTACCTTGG AAAAGGCGTG TTAACAGCTG TTAACAACGT AAACGAAATC	180
ATTGCTCCAG AGCTTCTTGG CTTTGATGTA	210

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ser	Arg	Gly	Asn	Pro	Thr	Leu	Glu	Val	Glu	Val	Tyr	Thr	Glu	Ser	Gly
1															15

Ala	Phe	Gly	Arg	Gly	Met	Val	Pro	Ser	Gly	Ala	Ser	Thr	Gly	Glu	His
															30

Glu	Ala	Val	Glu	Leu	Arg	Asp	Gly	Asp	Lys	Ser	Arg	Tyr	Gly	Gly	Leu
															45

Gly	Thr	Gln	Lys	Ala	Val	Asp	Asn	Val	Asn	Ile	Ile	Ala	Glu	Ala	
															60

Ile	Ile	Gly	Tyr	Asp	Val

65 70

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *B. subtilis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ser Arg Gly Asn Pro Thr Val Glu Val Glu Val Tyr Thr Glu Thr Gly
1 5 10 15

Ala Phe Gly Arg Ala Leu Val Pro Ser Gly Ala Ser Thr Gly Gln Tyr
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Asp Arg Tyr Leu Gly Lys
35 40 45

Gly Val Leu Thr Ala Val Asn Asn Val Asn Glu Ile Ile Ala Pro Glu
50 55 60

Leu Leu Gly Phe Asp Val
65 70

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1511 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *B. subtilis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCCCGCGGCA ACCCAACACT TGAAGTAGAA GTTTACACTG AATCAGGTGC TTTGGACGT	60
GGTATGGTTC CATCAGGAGC TTCTACTGGT GAACACGAAG CAGTTGAAC TCGCGACGGT	120
GACAAATCTC GTTACGGTGG TCTTGGTACA CAAAAAGCTG TTGACAACGT AAACAACATC	180
ATTGCTGATT CTATCATTGG CTACGATGTA CGTGATCAAC AAGCTATTGA CCGTGCTATG	240
ATCGCACTTG ACGGTACTCC TAACAAAGGT AAATTGGGTG CGAATGCAAT CCTCGGTGTG	300
TCTATCGCTG TAGCTCGTGC TGCTGCTGAC TACCTTGAAA TCCCACTTA CAGCTACCTT	360
GGTGGATTCA ACACTAAAGT TCTTCCAACT CCAATGATGA ACATCATCAA CGGTGGTTCT	420
CACTCTGACG CTCCAATCGC TTTCCAAGAG TTCATGATCT TGCCAGTTGG TGCGCCAACA	480
TTTAAAGAAA CCCTTCGTTA CGGTGCTGAA ATCTTCCACG CTCTTAAGAA AATCCTTAAA	540
TCACGTGGTT TGGAAACTGC CGTAGGTGAC GAAGGTGGAT TCGCTCCTCG TTTCGAAGGA	600
ACTGAAGATG GTGTTGAAAC TATCCTTGCT GCGATTGAAG CTGCTGGATA TGTACCAGGT	660

AAAGACGTAT TTCTCGGATT TGACTGTGCT TCATCAGAAT TCTACGATAA AGAACGTAAA	720
GTTTACGACT ACAC TAAATT TGAAGGC GAA GGTGCTGCTG TT CGTACATC TGCAGAACAA	780
ATCGACTACC TTGAAGAATT GGTAAACAAA TACCCAATCA TCACTATTGA AGATGGTATG	840
GATGAAAACG ACTGGGATGG TTGGAAAGCT CTTACTGAAC GTCTGGTAA GAAAGTACAA	900
CTTGTGGTG ACGACTTCTT CGTAACAAAC ACTGACTACC TTGCACGTGG TATCCAAAAAA	960
GGTGCTGCTA ACTCAATCCT TATCAAAGTT AACCAAATCG GTACTCTTAC TGAAACTTTT	1020
GAAGCTATCG AAATGGCTAA AGAAGCTGGT TACACTGCTG TTGTATCACA CCGTTCAGGT	1080
GAAACTGAAG ATTCAACAAT CGCTGATATC GCGATTGCAA CTAACGCAGG ACAAAATCAAG	1140
ACTGGTCAC TTTCACGTAC AGACCGTATC GCTAAATACA ACCAATTGCT TCGTATCGAA	1200
GACCAACTTG GTGAAGTAGC AGAATATCGT GGATTGAAAT CATTCTACGA CCTTAAAAAA	1260
TAAAATAGTA CAGTGAACTA TTTTATCCCC GAACCATGAA ATTCAAATTC CGGCCCTTGA	1320
TTTAACAACT TTCTAGCCCC CCGATTATC CGAGGGGATA TTTTGTTTT AAAGGGCAA	1380
AAGGACAAAC TTTTAAAT CTTTCATAC ATTACCATAC TTGATGCTCC CCTCCTTCTT	1440
GTACTGGTTT AAGAATATAC TCCATCGATG ACACCGTCGG ATGTTTAGGA TTATTCGACA	1500
ACAAGTTGAA T	1511

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 420 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *B. subtilis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ser Arg Gly Asn Pro Thr Leu Glu Val Glu Val Tyr Thr Glu Ser Gly
1 5 10 15

Ala Phe Gly Arg Gly Met Val Pro Ser Gly Ala Ser Thr Gly Glu His
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Ser Arg Tyr Gly Gly Leu
35 40 45

Gly Thr Gln Lys Ala Val Asp Asn Val Asn Asn Ile Ile Ala Asp Ser
50 55 60

Ile Ile Gly Tyr Asp Val Arg Asp Gln Gln Ala Ile Asp Arg Ala Met
65 70 75 80

Ile Ala Leu Asp Gly Thr Pro Asn Lys Gly Lys Leu Gly Ala Asn Ala
85 90 95

Ile Leu Gly Val Ser Ile Ala Val Ala Arg Ala Ala Asp Tyr Leu
100 105 110

Glu Ile Pro Leu Tyr Ser Tyr Leu Gly Gly Phe Asn Thr Lys Val Leu
115 120 125

Pro Thr Pro Met Met Asn Ile Ile Asn Gly Gly Ser His Ser Asp Ala
130 135 140

Pro Ile Ala Phe Gln Glu Phe Met Ile Leu Pro Val Gly Ala Pro Thr
145 150 155 160

Phe Lys Glu Thr Leu Arg Tyr Gly Ala Glu Ile Phe His Ala Leu Lys
165 170 175

Lys Ile Leu Lys Ser Arg Gly Leu Glu Thr Ala Val Gly Asp Glu Gly
180 185 190

Gly Phe Ala Pro Arg Phe Glu Gly Thr Glu Asp Gly Val Glu Thr Ile
195 200 205

Leu Ala Ala Ile Glu Ala Ala Gly Tyr Val Pro Gly Lys Asp Val Phe
210 215 220

Leu Gly Phe Asp Cys Ala Ser Ser Glu Phe Tyr Asp Lys Glu Arg Lys
225 230 235 240

Val Tyr Asp Tyr Thr Lys Phe Glu Gly Glu Gly Ala Ala Val Arg Thr
245 250 255

Ser Ala Glu Gln Ile Asp Tyr Leu Glu Glu Leu Val Asn Lys Tyr Pro
260 265 270

Ile Ile Thr Ile Glu Asp Gly Met Asp Glu Asn Asp Trp Asp Gly Trp
275 280 285

Lys Ala Leu Thr Glu Arg Leu Gly Lys Lys Val Gln Leu Val Gly Asp
290 295 300

Asp Phe Phe Val Thr Asn Thr Asp Tyr Leu Ala Arg Gly Ile Gln Lys
305 310 315 320

Gly Ala Ala Asn Ser Ile Leu Ile Lys Val Asn Gln Ile Gly Thr Leu

325

330

335

Thr Glu Thr Phe Glu Ala Ile Glu Met Ala Lys Glu Ala Gly Tyr Thr
340 345 350

Ala Val Val Ser His Arg Ser Gly Glu Thr Glu Asp Ser Thr Ile Ala
355 360 365

Asp Ile Ala Ile Ala Thr Asn Ala Gly Gln Ile Lys Thr Gly Ser Leu
370 375 380

Ser Arg Thr Asp Arg Ile Ala Lys Tyr Asn Gln Leu Leu Arg Ile Glu
385 390 395 400

Asp Gln Leu Gly Glu Val Ala Glu Tyr Arg Gly Leu Lys Ser Phe Tyr
405 410 415

Asp Leu Lys Lys
420

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 429 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AAAAAAGGTA GCAGAAGCTG AGAAGAAGGT TGAAGAAGCT GAGAAAAAAG CCAAGGATCA	60
AAAAGAAGAA GATGCCGTA ACTACCCAAC CAATACTTAC AAAACGCTTG ACCTTGAAAT	120
TGCTGAGTCC GATGTGAAAG TTAAAGAACG GGAGCTTGAA CTAGTAAAAG AGGAAGCTAA	180
GGAACCTCGA GACGAGGAAA AAATTAAGCA AGCAAAAGCG AAAGTTGAGA GTAAAAAAGC	240
TGAGGCTACA AGGTTAGAAA ACATCAAGAC AGATCGTAA AAAGCAGAAG AAGAAGCTAA	300
ACGAAAAGCA GCAGAAGAAG ATAAAGTTAA AGAAAAACCA GCTGAACAAAC CACAACCAGC	360
GCCGGTTACT CAACCAGAAA AACCAAGCTCC AAAACCAGAG AAGCCAGCTG AACAAACAAA	420
AGCAGAAAA	429

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Lys Lys Val Ala Glu Ala Glu Lys Lys Val Glu Glu Ala Glu Lys Lys
1 5 10 15

Ala Lys Asp Gln Lys Glu Glu Asp Arg Arg Asn Tyr Pro Thr Asn Thr
20 25 30

Tyr Lys Thr Leu Asp Leu Glu Ile Ala Glu Ser Asp Val Lys Val Lys
35 40 45

Glu Ala Glu Leu Glu Leu Val Lys Glu Glu Ala Lys Glu Pro Arg Asp
50 55 60

Glu Glu Lys Ile Lys Gln Ala Lys Ala Lys Val Glu Ser Lys Lys Ala
65 70 75 80

Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp Arg Lys Lys Ala Glu
85 90 95

Glu Glu Ala Lys Arg Lys Ala Ala Glu Glu Asp Lys Val Lys Glu Lys
100 105 110

Pro Ala Glu Gln Pro Gln Pro Ala Pro Val Thr Gln Pro Glu Lys Pro
115 120 125

Ala Pro Lys Pro Glu Lys Pro Ala Glu Gln Pro Lys Ala Glu
130 135 140

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTCTTNCTC GATGTCTCNG TNGCCAT

27

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AGCATAGTCT TCTTCGACTT GTTGATCATC

30

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2462 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCAAGCAGCC ACTTCTTCTA ATATGGAAA GACAGAACAT AGGAAAGCTG CTAAACAAGC	60
GTCGATGAAT ATATAGAAAA AATGATGAGG GAGATTCAAC TAGATAGAAG AAAACATACC	120
CAAAATGTCG CCTTAAACAT AAAGTTGAGC GCAATTAAAA CGAAGTATTG GCGTGAATTA	180
AATGTTTAG AAGAGAAGTC GAAAGATGAG TTGCCGTCAG AAATAAAAGC AAAGTTAGAC	240
GCAGCTTTG AGAAGTTAA AAAAGATACA TTGAAACCAG GAGAAAAGGT AGCAGAAGCT	300
AAGAAGAAGG TTGAAGAACG TAAGAAAAAA GCCGAGGATC AAAAAGAAGA AGATCGTCGT	360
AACTACCAA CCAACTTA CAAAACGCTT GAACTTGAAA TTGCTGAGTT CGATGTGAAA	420
GTAAAGAAG CGGAGCTTGA ACTAGTAAAA GAGGAAGCTA AAGAATCTCG AAACGAGGGC	480
ACAATTAAGC AAGCAAAAGA GAAAGTTGAG AGTAAAAAAAG CTGAGGCTAC AAGGTTAGAA	540
AACATCAAGA CAGATCGTAA AAAAGCAGAA GAAGAAGCTA AACGAAAAGC AGATGCTAAG	600
TTGAAGGAAG CTAATGTAGC GACTTCAGAT CAAGGTAAAC CAAAGGGGCG GGCAAAACGA	660

GGAGTTCCTG GAGAGCTAGC AACACCTGAT AAAAAAGAAA ATGATGCGAA GTCTTCAGAT	720
TCTAGCGTAG GTGAAGAAC TCTTCCAAGC TCATCCCTGA AATCAGGAAA AAAGGTAGCA	780
GAAGCTGAGA AGAAGGTTGA AGAAGCTGAG AAAAAAGCCA AGGATCAAAA AGAAGAAGAT	840
CGCCGTAAC TACCAACCAA TACTTACAAA ACGCTTGACC TTGAAATTGC TGAGTCCGAT	900
GTGAAAGTTA AAGAAGCGGA GCTTGAAC TA GTAAAAGAGG AAGCTAAGGA ACCTCGAGAC	960
GAGGAAAAAA TTAAGCAAGC AAAAGCGAAA GTTGAGAGTA AAAAAGCTGA GGCTACAAGG	1020
TTAGAAAACA TCAAGACAGA TCGTAAAAAA GCAGAAGAAG AAGCTAAACG AAAAGCAGCA	1080
GAAGAAGATA AAGTTAAAGA AAAACCAGCT GAACAACCAC AACCAAGGCC GGCTACTCAA	1140
CCAGAAAAAC CAGCTCCAAA ACCAGAGAAG CCAGCTGAAC AACCAAAAGC AGAAAAAAACA	1200
GATGATCAAC AAGCTGAAGA AGACTATGCT CGTAGATCAG AAGAAGAATA TAATCGCTTG	1260
ACTCAACAGC AACCGCCAAA AACTGAAAAA CCAGCACAC CATCTACTCC AAAAACAGGC	1320
TGGAAACAAG AAAACGGTAT GTGGTACTTC TACAATACTG ATGGTTCAAT GGCAACAGGA	1380
TGGCTCCAAA ACAACGGTTC ATGGTACTAT CTAAACGCTA ATGGTGCTAT GGCGACAGGA	1440
TGGCTCCAAA ACAATGGTTC ATGGTACTAT CTAAACGCTA ATGGTTCAAT GGCAACAGGA	1500
TGGCTCCAAA ACAATGGTTC ATGGTACTAC CTAAACGCTA ATGGTGCTAT GGCGACAGGA	1560
TGGCTCCAAT ACAATGGTTC ATGGTACTAC CTAAACAGCA ATGGCGCTAT GGCGACAGGA	1620
TGGCTCCAAT ACAATGGCTC ATGGTACTAC CTCAACGCTA ATGGTGATAT GGCGACAGGA	1680
TGGCTCCAAA ACAACGGTTC ATGGTACTAC CTCAACGCTA ATGGTGATAT GGCGACAGGA	1740
TGGCTCCAAT ACAACGGTTC ATGGTATTAC CTCAACGCTA ATGGTGATAT GGCGACAGGT	1800

TGGGTGAAAG ATGGAGATACTGGTACTAT CTTGAAGCAT CAGGTGCTAT GAAAGCAAGC	1860
CAATGGTCA AAGTATCAGA TAAATGGTAC TATGTCAATG GCTCAGGTGC CCTTGCAGTC	1920
AACACAACTG TAGATGGCTA TGGAGTCAAT GCCAATGGTG AATGGGTAAA CTAAACCTAA	1980
TATAACTAGT TAATACTGAC TTCCTGTAAG AACTTTTAA AGTATTCCCT ACAAATACCA	2040
TATCCTTCA GTAGATAATA TACCCTTGTA GGAAGTTAG ATTAAAAAT AACTCTGTAA	2100
TCTCTAGCCG GATTATAGC GCTAGAGACT ACGGAGTTTT TTTGATGAGG AAAGAATGGC	2160
GGCATTCAAG AGACTCTTA AGAGAGTTAC GGGTTTAAA CTATTAAGCC TTCTCCAATT	2220
GCAAGAGGCT TCAATCTCTG CTAGGGTGCT AGCTTGCAGA ATGGCTCCAC GGAGTTGGC	2280
AGCGCCAGAT GTTCCACGGA GATAGTGAGG AGCGAGGCCG CGGAATTACAC GAACTGCGAC	2340
GTTTCTCCT TTGAGGTTAA TCAATCGTT CAAGTGGTCG TAGGCGATCT TCATCTTGTC	2400
TTCAAAGGTC AAATCAGGTA GGATTTCTCC TGTTCAAAG TTATGGTGG CCCTGGTTGA	2460
AG	2462

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 631 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met Met Arg Glu Ile Gln Leu Asp Arg Arg Lys His Thr Gln Asn Val
1 5 10 15

Ala Leu Asn Ile Lys Leu Ser Ala Ile Lys Thr Lys Tyr Leu Arg Glu
20 25 30

Leu Asn Val Leu Glu Glu Lys Ser Lys Asp Glu Leu Pro Ser Glu Ile
35 40 45

Lys Ala Lys Leu Asp Ala Ala Phe Glu Lys Phe Lys Lys Asp Thr Leu
50 55 60

Lys Pro Gly Glu Lys Val Ala Glu Ala Lys Lys Lys Val Glu Glu Ala
65 70 75 80

Lys Lys Lys Ala Glu Asp Gln Lys Glu Glu Asp Arg Arg Asn Tyr Pro
85 90 95

Thr Asn Thr Tyr Lys Thr Leu Glu Leu Glu Ile Ala Glu Phe Asp Val
100 105 110

Lys Val Lys Glu Ala Glu Leu Glu Leu Val Lys Glu Glu Ala Lys Glu
115 120 125

Ser Arg Asn Glu Gly Thr Ile Lys Gln Ala Lys Glu Lys Val Glu Ser
130 135 140

Lys Lys Ala Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp Arg Lys
145 150 155 160

Lys Ala Glu Glu Ala Lys Arg Lys Ala Asp Ala Lys Leu Lys Glu
165 170 175

Ala Asn Val Ala Thr Ser Asp Gln Gly Lys Pro Lys Gly Arg Ala Lys
 180 185 190

Arg Gly Val Pro Gly Glu Leu Ala Thr Pro Asp Lys Lys Glu Asn Asp
 195 200 205

Ala Lys Ser Ser Asp Ser Ser Val Gly Glu Glu Thr Leu Pro Ser Ser
 210 215 220

Ser Leu Lys Ser Gly Lys Lys Val Ala Glu Ala Glu Lys Lys Val Glu
 225 230 235 240

Glu Ala Glu Lys Lys Ala Lys Asp Gln Lys Glu Glu Asp Arg Arg Asn
 245 250 255

Tyr Pro Thr Asn Thr Tyr Lys Thr Leu Asp Leu Glu Ile Ala Glu Ser
 260 265 270

Asp Val Lys Val Lys Glu Ala Glu Leu Glu Leu Val Lys Glu Glu Ala
 275 280 285

Lys Glu Pro Arg Asp Glu Glu Lys Ile Lys Gln Ala Lys Ala Lys Val
 290 295 300

Glu Ser Lys Lys Ala Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp
 305 310 315 320

Arg Lys Lys Ala Glu Glu Ala Lys Arg Lys Ala Ala Glu Glu Asp
 325 330 335

Lys Val Lys Glu Lys Pro Ala Glu Gln Pro Gln Pro Ala Pro Ala Thr
 340 345 350

Gln Pro Glu Lys Pro Ala Pro Lys Pro Glu Lys Pro Ala Glu Gln Pro
 355 360 365

Lys Ala Glu Lys Thr Asp Asp Gln Gln Ala Glu Glu Asp Tyr Ala Arg

370

375

380

Arg Ser Glu Glu Glu Tyr Asn Arg Leu Thr Gln Gln Gln Pro Pro Lys
385 390 395 400

Thr Glu Lys Pro Ala Gln Pro Ser Thr Pro Lys Thr Gly Trp Lys Gln
405 410 415

Glu Asn Gly Met Trp Tyr Phe Tyr Asn Thr Asp Gly Ser Met Ala Thr
420 425 430

Gly Trp Leu Gln Asn Asn Gly Ser Trp Tyr Tyr Leu Asn Ala Asn Gly
435 440 445

Ala Met Ala Thr Gly Trp Leu Gln Asn Asn Gly Ser Trp Tyr Tyr Leu
450 455 460

Asn Ala Asn Gly Ser Met Ala Thr Gly Trp Leu Gln Asn Asn Gly Ser
465 470 475 480

Trp Tyr Tyr Leu Asn Ala Asn Gly Ala Met Ala Thr Gly Trp Leu Gln
485 490 495

Tyr Asn Gly Ser Trp Tyr Tyr Leu Asn Ser Asn Gly Ala Met Ala Thr
500 505 510

Gly Trp Leu Gln Tyr Asn Gly Ser Trp Tyr Tyr Leu Asn Ala Asn Gly
515 520 525

Asp Met Ala Thr Gly Trp Leu Gln Asn Asn Gly Ser Trp Tyr Tyr Leu
530 535 540

Asn Ala Asn Gly Asp Met Ala Thr Gly Trp Leu Gln Tyr Asn Gly Ser
545 550 555 560

Trp Tyr Tyr Leu Asn Ala Asn Gly Asp Met Ala Thr Gly Trp Val Lys
565 570 575

Asp Gly Asp Thr Trp Tyr Tyr Leu Glu Ala Ser Gly Ala Met Lys Ala
580 585 590

Ser Gln Trp Phe Lys Val Ser Asp Lys Trp Tyr Tyr Val Asn Gly Ser
595 600 605

Gly Ala Leu Ala Val Asn Thr Thr Val Asp Gly Tyr Gly Val Asn Ala
610 615 620

Asn Gly Glu Trp Val Asn Glx
625 630